Encoding variability accounts for false recognition of noncorresponding lures on the forced-choice Mnemonic Similarity Task

Leslie Rollins, Alexis Khuu, and Nafeesa Lodi

Department of Psychology, Christopher Newport University, Newport News, Virginia 23606, USA

On forced-choice tests of recognition memory, performance is best when targets are paired with novel foils (A-X), followed by corresponding lures (A-A′), and then noncorresponding lures (A-B′). The current study tested the prediction that encoding variability accounts for reduced performance on A-B′ trials. Young adults ($n = 43$) completed the forced-choice Mnemonic Similarity Task while eye movements were recorded. Errors on the A-B′ test format were attributable to encoding variability; incorrect A-B′ trials were associated with more fixations to the B stimulus than the A stimulus at encoding. However, encoding variability did not account for accuracy on the A-A′ test format.

Pattern separation is a computational process performed by the hippocampus to support the formation of distinct memory representations by orthogonalizing overlapping input (e.g., Treves and Rolls 1992; Norman and O'Reilly 2003). Pattern separation can be taxed in humans using the Mnemonic Similarity Task (MST), which requires differentiating between previously viewed objects (i.e., targets), perceptually similar objects (i.e., lures), and novel objects (i.e., foils; Kirwan and Stark 2007). Huffman and Stark (2017) recently developed a two-alternative forced-choice version of the MST. Following incidental encoding, participants identified which of two objects was previously encountered. Targets were either paired with a foil (i.e., A-X), a corresponding lure (i.e., A-A′), or a noncorresponding lure (i.e., A-B′), similar to previous research (e.g., Hintzman 1988; Jeneson et al. 2010; Migo et al. 2014). Performance was highest for the A-X test format followed by the A-A′ test format and then the A-B′ test format. The global matching model MINERVA 2 (Hintzman 1984, 1988) suggested that encoding variability could partially account for poorer performance for the A-B′ test format relative to the A-A′ test format (Huffman and Stark 2017). Specifically, false recognition of the noncorresponding lure may be due to superior encoding of the B stimulus relative to the A stimulus, resulting in a stronger global match for the B′ stimulus than the target (Huffman and Stark 2017). The primary aim of the current study was to test this prediction by using visual processing to index encoding quality.

Visual processing can provide important insights into memory (e.g., Hannula et al. 2010). Consistent with foveated regions being processed with higher visual acuity (Pertzov et al. 2009), subsequent memory is enhanced for stimuli associated with higher fixation counts and more distributed fixations at encoding (Kafkas and Montaldi 2011; Damiano and Walther 2019). Further, restricting eye movements at encoding leads to recognition memory deficits (Henderson et al. 2005; Damiano and Walther 2019). Molitor et al. (2014) recently examined eye movements while young adults performed the continuous recognition version of the MST. Lure false alarms (i.e., lures identified as "old") were associated with fewer fixations during first presentation relative to lure correct rejections (i.e., lures identified as "similar"). This finding suggested that encoding may partially account for deficits in mnemonic discrimination.

The present study tested the prediction that encoding variability accounts for performance on the forced-choice MST by examining visual processing at encoding. Performance was hypothesized to be highest for the A-X test format followed by the A-A′ test format and then the A-B′ test format (Huffman and Stark 2017). On A-X test trials, the memory signals generated by the target and novel foil are uncorrelated, and the target should elicit the retrieval of a stronger memory representation than the novel foil. For the A-A′ test format, the test probes generate correlated memory signals due to featural similarity. However, the memory signal should be stronger for A than A′ because it shares more features with the encoded item (Hintzman 1988). We anticipated that fixations at encoding, which may index successful feature encoding, would influence accuracy on A-A′ test trials. Based on the results observed by Molitor et al. (2014), we hypothesized that higher fixation counts at encoding would be observed for correct than incorrect A-A′ test trials. The A-B′ test format is theorized to be more challenging than the A-A′ test format because each test probe activates the retrieval of distinct memory representations (Hintzman 1988). Thus, the recognition memory judgment is based on the relative strength of the representations elicited by the test probes. If more features were encoded for the B stimulus than the A stimulus, the B′ stimulus may elicit a stronger memory signal than the target stimulus, resulting in false recognition. Consistent with the MINERVA 2 model and the proposal by Huffman and Stark (2017), we hypothesized that incorrect A-B′ trials would be associated with higher fixation counts at encoding to the B stimulus than the A stimulus. Correct A-B′ trials may either be associated with enhanced encoding of the A stimulus or equivalent encoding of the A and B stimuli. When the encoding quality is similar, the target stimulus (i.e., A) generates a stronger memory signal than the lure (i.e., B′).

Forty-three young adults ($M = 19.71 \pm 1.34$ yr; 35 females, 8 males) were included in the study. Six additional participants completed the study but were excluded due to technical error ($n = 4$), experimenter error ($n = 1$), or failure to meet inclusion criteria ($n = 1$).

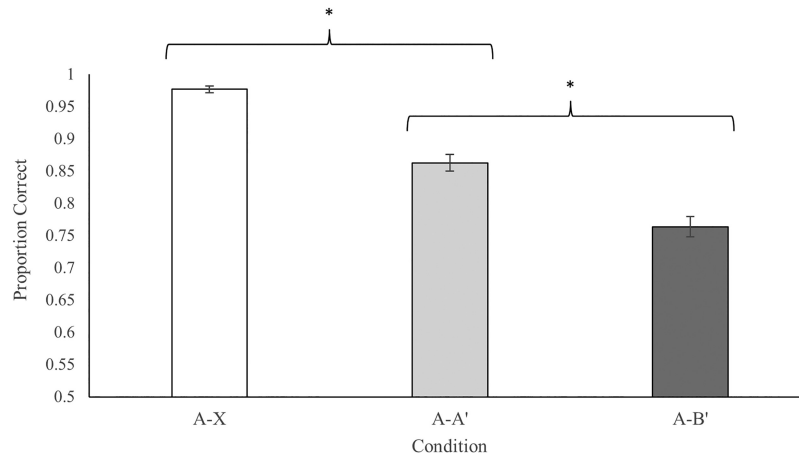**Corresponding author: leslie.rollins@cnu.edu**

Figure 1. Memory performance in each test format of the forced-choice MST. Error bars reflect standard errors. * indicates $p < 0.05$.

Participants provided informed consent and then were seated in front of a 24 inch monitor equipped with a Tobii X3-120 eye-tracking system (Tobii Technology, Danderyd, Sweden). Tobii Studio Pro software presented stimuli and collected eye-tracking and behavioral data. Before initiating the experiment, participants completed a standard nine-point calibration procedure.

The forced-choice MST was modeled after Huffman and Stark (2017). A total of 125 pairs of pictures were selected from the Stark Laboratory's database of stimuli for the MST (https://faculty.sites.uci.edu/starklab/mnemonic-similarity-task-mst/). Stimuli were divided into five sets of 25 images evenly distributed across five levels of mnemonic similarity established in previous research (e.g., Stark et al. 2013). The stimulus sets were used to counterbalance images identified as targets, lures, and foils across participants. Participants incidentally encoded 100 pictures by making an indoor/outdoor judgment. Each stimulus was presented for 2000 msec with a 500 msec interstimulus interval. During retrieval, participants completed a two-alternative forced-choice memory test, which included three test formats; targets were either presented with a corresponding lure, (A-A'), noncorresponding lure (A-B'), or foil (A-X). Each test format included 25 trials, and trials were presented in a random order. For each trial, participants identified which of the two stimuli they previously viewed. Each pair remained on the screen until the participants made their decision. Trials for each test format were evenly distributed among the five mnemonic similarity levels, and stimuli for the A-B' test format were matched by mnemonic similarity level to account for retrieval difficulty.

Fixation counts during encoding and retrieval were extracted for an area of interest that encompassed each picture (550 × 550 pixels). A fixation was defined as a series of data points in which the eyes did not shift more than 35 pixels for at least 60 msec. Data were sorted as a function of test format (A-A', A-B', A-X), stimulus, and response accuracy (correct, incorrect). Participants were required to have a minimum of four trials per condition to be included in the analyses of eye movements (Hannula et al. 2010; Molitor et al. 2014). This criterion resulted in the inclusion of 35 partici-

pants for the analysis of the A-B' test format and 25 participants for the analysis of the A-A' test format. Performance was too high in the A-X condition to assess eye-tracking data for correct versus incorrect trials. Eye-tracking data from retrieval did not provide insight into mnemonic processes; participants fixated more on the stimulus they selected, regardless of response accuracy (analyses available upon request).

Performance was above chance (0.5) for all test formats, $ts_{(42)} = 16.52$–91.77, $Ps < 0.01$. However, mnemonic discrimination varied among test formats, $F_{(2,84)} = 103.88$, $P < 0.01$; recognition was the best for the A-X test format followed by the A-A' test format and then the A-B' test format (see Fig. 1). A 2 Test Format (A-A', A-B') × 5 Mnemonic Similarity (L1: Most Similar, L2, L3, L4, L5: Least Similar) repeated-measures ANOVA assessed the effect of mnemonic similarity on performance. There were significant main effects of Test Format, $F_{(1,42)} = 28.28$, $P < 0.01$, and Mnemonic Similarity, $F_{(4,168)} = 20.63$, $P < 0.01$. Targets were better identified when paired with Level 4 and Level 5 Lures, which were more mnemonically distinct from targets, than Level 1–3 Lures (see Fig. 2). The Test Format × Mnemonic Similarity interaction was not significant, $F_{(4,168)} = 2.04$, $P = 0.09$.

The primary aim of the present study was to determine whether reduced accuracy on the A-B' test format was attributable to encoding variability. Mean fixation counts from encoding were examined using a 2 Stimulus (A, B) × 2 Subsequent Accuracy (Correct, Incorrect) repeated-measures ANOVA. Consistent with the encoding variability hypothesis, the analysis revealed a significant Stimulus × Subsequent Accuracy interaction, $F_{(1,34)} = 7.34$, $P = 0.01$. Follow-up paired sample t-tests showed that incorrect A-B' trials were associated with more fixations to the B stimulus than the A stimulus at encoding, $t_{(34)} = -2.24$, $P = 0.02$ (see Fig. 3). The difference in fixation counts at encoding when participants correctly chose the target of an A-B' trial was not significant, $t_{(34)} = 1.41$, $P = 0.08$, although there was a trend for preferential viewing of the A stimulus. We had hypothesized that fixation counts at encoding would be lower for incorrect A-A' trials than correct A-A' trials.
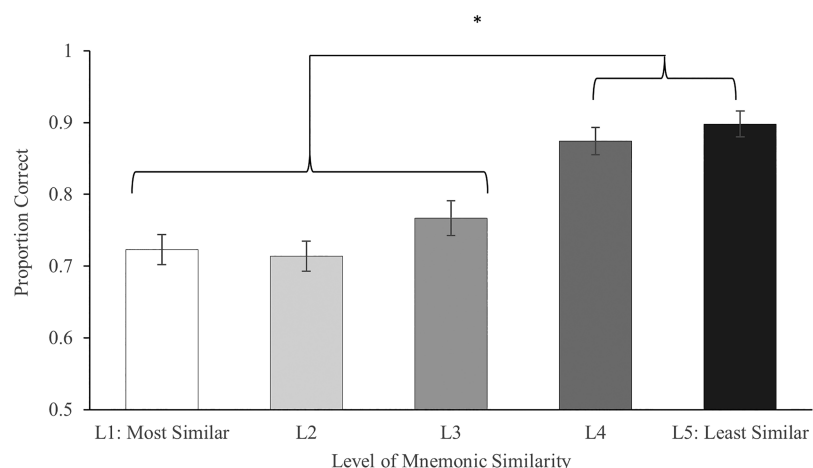

Figure 2. Effect of mnemonic similarity on memory performance. Error bars reflect standard errors. * indicates $p < 0.05$.
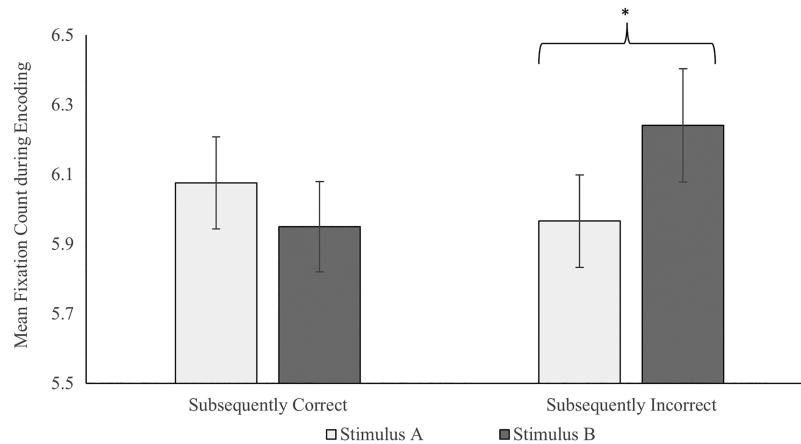
Figure 3. Mean fixation counts at encoding for stimuli subsequently included in the A-B′ test format. Subsequently incorrect trials were associated with more fixations to the original version of the lure (*B*) than the target stimulus (*A*) at encoding. Fixation counts at encoding for the A and B stimuli did not differ for subsequently correct trials. Error bars reflect standard errors. * indicates $p < 0.05$.

However, fixations at encoding were comparable regardless of subsequent accuracy on the A-A′ test format trials, $t_{(24)} = 1.04$, $P = 0.15$.

Performance on the forced-choice MST was best for the A-X test format, followed by the A-A′ test format, and then the A-B′ test format. This pattern of results aligns with previous studies (Hintzman 1988; Jeneson et al. 2010; Huffman and Stark 2017). Multiple factors contribute to the cognitive demand associated with the A-B′ test format relative to the A-A′ test format. A-A′ test trials elicit the retrieval of correlated memory signals; however, the strength of the retrieved signal should be stronger for the target than corresponding lure due to more features being shared with the encoded item. In contrast, probes on A-B′ test trials generate the retrieval of uncorrelated memory signals, which leads to interference and encoding variability enhancing the difficulty of A-B′ test trials (Huffman and Stark 2017). Interference refers to the probability of the lure resembling other encountered items, and encoding variability refers to variability in the number of features encoded for each stimulus. Huffman and Stark (2017) conducted simulations of the MINERVA 2 model to assess how removing encoding variability influenced performance on the A-A′, A-B′, and A-X test formats. Removing encoding variability substantially increased performance on the A-B′ test format; however, only minor improvements in performance were observed for the A-A′ and A-X test formats. Based on these simulations, Huffman and Stark (2017) argued that errors on the A-B′ test format are likely partially attributable to the superior encoding of the B stimulus relative to the A stimulus, which results in a stronger global match to the lure (B′) relative to the target during the forced-choice recognition trial.

The current study empirically evaluated the encoding variability hypothesis by examining visual processing at encoding. Errors on the A-B′ test format were associated with more fixations to the B stimulus than the A stimulus at encoding. Therefore, encoding variability partially accounts for reduced performance on the A-B′ test format relative to the A-A′ and A-X test formats. For the A-X and A-B′ test formats, uncorrelated memory signals are retrieved for each probe, and the recognition memory judgment is based on the relative strength of those signals. Performance is high on the A-X test format because the strength of the memory signal associated with the foil should be low. In contrast, the strength of the retrieved memory signals should be more comparable for an A-B′ test trial. The present data suggest that false recognition of a noncorresponding lure (i.e., B′) can be explained by more effective encoding of the original version of that stimulus (i.e., B) than the

target (i.e., A). There was also a trend for correct A-B′ responses to be associated with higher fixations to the target than the original version of the noncorresponding lure (i.e., B). Our data contradict a recall-to-reject account of performance on the A-B′ test format (Rotello and Heit 1999, 2000). Recall-to-reject processing would involve the recollection of features associated with the originally encoded B stimulus in order to reject the lure (i.e., B′) and accurately select the target (i.e., A). If recall-to-reject processing was being recruited, fixation counts would have been higher for the B stimulus than the A stimulus for correct A-B′ trials.

Using the continuous recognition version of the MST, Molitor et al. (2014) observed that lure false alarms were associated with fewer fixations than hits and lure correct rejections at first presentation. Similarly, a MINERVA 2 simulation by Huffman and Stark (2017) showed that reducing the encoding parameter, which reflected the probability of each feature being encoded, reduced performance on the A-A′ test format but not the A-B′ or the A-X test format. Based on this evidence, we expected fixations at encoding to be indicative of accuracy on A-A′ trials. However, this effect was not statistically significant in the present study. The version of the MST used for each study likely explains why the present results differ from those observed by Molitor et al. (2014). The continuous recognition version of the task used by Molitor et al. (2014) was dependent on the successful encoding of features that would subsequently vary between the lure and the encoded item as well as the cued retrieval of the encoded representation by the lure during the test trial. Because the A-A′ test format simultaneously presents the target and corresponding lure, the effect size associated with the overall probability of feature encoding on mnemonic discrimination may be smaller in the forced-choice than the continuous recognition version of the MST.

Eye-tracking research can provide further insights into pattern separation and pattern completion. The present study and Molitor et al. (2014) examined the number of fixations elicited to the entire stimulus at encoding. Research is needed to determine whether precisely what features are encoded is more critical than the number of features encoded. Unfortunately, this hypothesis cannot be tested using the MST because items vary in overall mnemonic similarity rather than the manipulation of individual features (Hunsaker and Kesner 2013). Future studies could test the prediction that lure false alarms would be associated with fewer fixations to features that are subsequently manipulated. Such a result would be consistent with the claim that false alarms to lures may be partially attributable to deficient pattern separation during learning rather than wholly attributable to pattern completion at retrieval (Molitor et al. 2014). Additionally, an eye-tracking study would be informative to assess the claim that deficits in feature encoding account for the age-related decline in mnemonic discrimination in older adults (Huffman and Stark 2017).

In conclusion, the present study suggests that encoding variability accounts for false recognition of lures during the A-B′ test format of the forced-choice MST. Specifically, for incorrect A-B′ trials, fixation counts at encoding were higher for the B stimulus than the A stimulus. In contrast, encoding variability did not account for accuracy on the A-A′ test format. These findings suggest that mnemonic discrimination on the A-B′ test format of the forced-choice version of the MST is susceptible to

encoding variability. Future studies need to account for encoding variability to effectively examine processes that underlie mnemonic discrimination.

References

Damiano C, Walther DB. 2019. Distinct roles of eye movements during memory encoding and retrieval. *Cognition* **184:** 119–129. doi:10.1016/j.cognition.2018.12.014

Hannula DE, Althoff RR, Warren DE, Riggs L, Cohen NJ, Ryan JD. 2010. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Front Hum Neurosci* **4:** 166. doi:10.3389/fnhum.2010.00166

Henderson JM, Williams CC, Falk RJ. 2005. Eye movements are functional during face learning. *Mem Cognit* **33:** 98–106. doi:10.3758/BF03195300

Hintzman DL. 1984. MINERVA 2: a simulation model of human memory. *Behav Res Meth Instrum Comput* **16:** 96–101. doi:10.3758/BF03202365

Hintzman DL. 1988. Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychol Rev* **95:** 528–551. doi:10.1037/0033-295X.95.4.528

Huffman DJ, Stark CEL. 2017. Age-related impairment on a forced-choice version of the Mnemonic Similarity Task. *Behav Neurosci* **131:** 55–67. doi:10.1037/bne0000180

Hunsaker MR, Kesner RP. 2013. The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neurosci Behav Rev* **37:** 36–58. doi:10.1016/j.neubiorev.2012.09.014

Jeneson A, Kirwan CB, Hopkins RO, Wixted JT, Squire LR. 2010. Recognition memory and the hippocampus: a test of the hippocampal contribution to recollection and familiarity. *Learn Mem* **17:** 63–70. doi:10.1101/lm.1546110

Kafkas A, Montaldi D. 2011. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Q J Exp Psychol* **64:** 1971–1989. doi:10.1080/17470218.2011.588335

Kirwan CB, Stark CEL. 2007. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learn Mem* **14:** 625–633. doi:10.1101/lm.663507

Migo EM, Quamme JR, Holmes S, Bendell A, Norman KA, Mayes AR, Montaldi D. 2014. Individual differences in forced-choice recognition memory: partitioning contributions of recollection and familiarity. *Q J Exp Psychol* **67:** 2189–2206. doi:10.1080/17470218.2014.910240

Molitor RJ, Ko PC, Hussey EP, Ally BA. 2014. Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. *Hippocampus* **24:** 666–672. doi:10.1002/hipo.22256

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev* **110:** 611–646. doi:10.1037/0033-295X.110.4.611

Pertzov Y, Avidan G, Zohary E. 2009. Accumulation of visual information across multiple fixations. *J Vis* **9:** 1–12. doi:10.1167/9.10.2

Rotello CM, Heit E. 1999. Two-process models of recognition memory: evidence for recall-to-reject. *J Mem Lang* **40:** 432–453. doi:10.1006/jmla.1998.2623

Rotello CM, Heit E. 2000. Associative recognition: a case of recall-to-reject processing. *Mem Cognit* **28:** 907–922. doi:10.3758/BF03209339

Stark SM, Yassa MA, Lacy JW, Stark CEL. 2013. A task to assess behavioral pattern separation (BPS) in humans. *Neuropsychologia* **51:** 2442–2449. doi:10.1016/j.neuropsychologia.2012.12.014

Treves A, Rolls ET. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2:** 189–199. doi:10.1002/hipo.450020209